Determination of the Optimal Multi-physics Ensemble Configuration for Predicting California Renewable Generation

Lab Directed Research and Development Report
LDRD Project 11-ERD-076

Matthew Simpson
Vera Bulaevskaya
Pedro Sotorrio
Tom Edmunds
Lawrence Livermore National Laboratory
Livermore, CA 94550

July 10, 2014

Executive Summary

A new methodology for determining the optimal multi-physics ensemble configuration for predicting California wind and solar resources with the Weather Research and Forecast (WRF) atmospheric model is presented. We reduced the enormous number of possible WRF physics configurations to a manageable set by first sampling schemes from each physics category that utilize different parameterization approaches and demonstrate varying degrees of sophistication. Latin Hypercube Sampling (LHS) design was next used to further reduce the initial set of physics configurations to 420 ensemble members by sampling the physics scheme 6-dimensional space. A WRF test case was run with the 420-member ensemble to study the sensitivity of wind and solar predictions to each of the main WRF physics categories. Analysis of variance (ANOVA) on the WRF sensitivity predictions revealed that the physics categories with the greatest effect on wind predictions were the planetary boundary layer and the land surface model. Choice of the shortwave radiation model is the primary source for solar flux variability in the ensemble predictions. ANOVA results indicated that the number of WRF physics combinations could be further reduced to 54 members and still account for the majority of the spread in the ensemble wind and solar prediction distribution. Several forecast case studies with dynamically evolving wind and solar conditions spanning all seasons were run with the 54-member ensemble. An analysis of ensemble forecast error and spread from the case studies indicated that only 18 ensemble members were needed to generate accurate wind and solar forecasts over California.

1. Introduction

Forecasting renewable generation with a numerical weather prediction (NWP) model is inherently difficult due to the intermittent nature of wind and solar resources. The majority of NWP prediction errors associated with renewable generation forecasting can be attributed to uncertainties in the model initial / lateral boundary conditions, as well as imperfect model physics. One method to account for the uncertainty associated with wind and solar forecasting and to improve the prediction skill is to run a NWP model with a suite of varying physics configurations over the same forecast period. By doing so, a multi-physics ensemble is generated that is capable of sampling sources of model error and estimating forecast uncertainty (Hou 2001, Murphy 2004, Eckel 2005, Berner 2010, Hacker 2011). In addition, the multi-physics ensemble approach provides a probabilistic forecast of the future state of the atmosphere that is particularly beneficial when modeling dynamically evolving wind and solar resources. The rate of growth in model uncertainty has been shown to be better captured by a multi-physics ensemble approach versus a multi-initial conditions approach for short-range weather forecasting (Strensrud 2000). This finding makes the multi-physics ensemble philosophy ideal for renewable forecasting since the relevant prediction timeframe for most utility grid operations is usually one or two days.

The need for a multi-physics forecasting approach arises from the fact that no single physics configuration (i.e. a deterministic run) can always perform the best relative to other configurations over a wide range of atmospheric stabilities and phenomena. Also, a single physics configuration is incapable of sampling known sources of uncertainty associated with model sub-grid scale processes and is therefore unable to capture a probability distribution of atmospheric conditions at those scales.

Project Goal

The goal of this study is to determine the optimal multi-physics ensemble configuration for predicting California renewable generation. While running a large ensemble is likely to capture more sources of uncertainty, it can be computationally costly or even infeasible. An optimal ensemble in this context is therefore one that is small enough to be computationally manageable, but is very similar to a larger ensemble in terms of its prediction accuracy and ability to capture uncertainty.

Finding such an optimal multi-physics ensemble configuration for California renewable resources will have two major benefits. First, private industry weather forecasting groups can use the project results as a foundation for improving the accuracy of existing ensemble prediction systems. Secondly, research groups, such as those at national labs that perform large-scale renewable integration studies (e.g. Edmunds et al., 2013), can use the findings to make the best use of available computing resources. By reducing the number of multi-physics ensemble members, computing resources are freed up to sample additional known sources of uncertainty and therefore improve the confidence in research conclusions.

Methodology

The research methodology used to identify the optimal multi-physics ensemble members for California wind and solar generation consisted of three main steps. The first step involved reducing the large number of possible physics combinations to a more manageable set. The approach for reducing the initial set of physics combinations is discussed in Section 3 and

involved sampling across the sophistication of physics schemes and Latin hypercube sampling. The second step involved performing an analysis of variance (ANOVA) on predicted wind and solar meteorological data generated from a WRF forecast simulation with a large number of ensemble members. The ANOVA of the generated weather data isolated physics schemes that were the major contributors to the variability in the ensemble predictions, thus allowing for further reduction in the number of physics configurations to be included in the ensemble. Results of the ANOVA are provided in Section 3. This second step will hereafter be referred to as the *sensitivity test*. The third and final step of the methodology involved generating renewable forecasts for several case studies with dynamically evolving wind and cloud conditions using the reduced physics configurations. The case study results were then compared to synthetically generated wind and solar observations in order to calculate prediction skill for the physics configurations we considered. ANOVA of the forecast errors in the case studies then determined if the set of physics configurations could be further reduced without negatively impacting the ensemble prediction skill or spread. Results of the forecast case study analysis are discussed in Section 4.

Ideally, the analysis of this research would involve renewable power generation, but these data are difficult to obtain from site operators. As a result, in the absence of generation data, we focused on meteorology (i.e. wind speed and solar insolation) for this study. Also, we used model-generated synthetic weather observations to verify ensemble forecast results instead of onsite meteorological data. This is due to limited availability of wind and solar observations made at renewable generation sites.

2. Renewable Forecast Design

2.1 Atmospheric Model

The numerical weather prediction (NWP) model used to generate wind and solar forecasts in this study was the non-hydrostatic, fully compressible Weather Research and Forecast (WRF) model (Skamarock et al., 2008). Specifically, version 3.5.1 of the advanced research dynamical core of the WRF code was used. WRF is a state-of-the-science community-based atmospheric model deigned to address both operational weather forecasting and broad atmospheric research needs. WRF was developed via collaboration among numerous academic, research, and government organizations to streamline the transfer of atmospheric research findings to an operational capacity. The large set of available model physics schemes and data assimilation approaches coupled with efficient model nesting make WRF appropriate for performing simulations on scales of motion from tens of meters to thousands of kilometers. Source code for WRF is available for download by the public at no cost through a National Center for Atmospheric Research (NCAR) supported website[1].

Several features make WRF the ideal atmospheric model to use for investigating the optimal multi-physics ensemble configuration for California renewable generation prediction. WRF is currently widely used by both research and private weather forecasting sectors. Therefore, any 'best practices' concluded from this study can be utilized by an existing large user base. In addition, WRF has numerous schemes available for each of the main physics categories (e.g. land surface model, shortwave radiation, etc.), leading to a large number of working physics

---

[1] http://www2.mmm.ucar.edu/wrf/users/download/get_source.html

configurations, which is ideal for any ensemble-based study. Also, WRF source code has been developed and optimized to run on high-performance parallel computing resources, allowing for a large number of high-resolution ensemble forecasts to be performed in a reasonable timeframe.

Model Domain

A total of 4 model domains were used for the WRF atmospheric modeling, as shown in Figure 1. The outermost model domain, which is labeled D1, has a horizontal grid spacing of 27 km and covers the majority of the Western United States. Also, the size of domain 1 is sufficient to simulate large-scale weather systems as they propagate across California. Nested model domain 2 (D2) covers all of California with a horizontal grid spacing of 9 km. Nested WRF sibling domains 3 and 4 (labeled D3 and D4, respectively) both have a horizontal grid spacing of 3 km and were constructed to cover the significant wind resource regions of California (e.g. Tehachapi, Gorgonio, and Altamont). The fine-scale 3 km grid spacing of domains 3 and 4 was necessary to accurately simulate wind flow and small-scale atmospheric features present in complex terrain. It should be noted that domains 3 and 4 have sufficiently high resolution that no cumulus parameterization was used following WRF documentation.
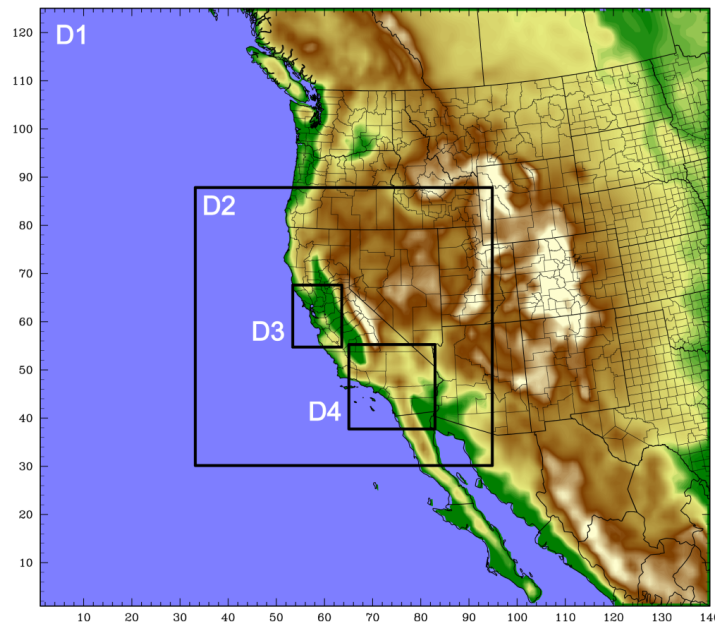

Figure 1. Nested WRF domain configuration used for the atmospheric ensemble modeling study. Model domain 1 (labeled D1) and domain 2 (D2) have a horizontal grid spacing of 27 and 9 km, respectively. Model domains 3 and 4 (D3, D4) both have a horizontal grid spacing of 3 km.

A total of 50 terrain-following vertical sigma levels were utilized for the WRF simulations. The sigma level distribution was designed to generate a vertical resolution of approximately 15 - 20 m in the lowest 200 m of the atmosphere. High vertical resolution was necessary in the surface layer for this study since we were running WRF in complex terrain and hub height wind speeds were a major focus of our analysis. Above 200 m, the sigma level vertical resolution was gradually stretched up to the model grid top at 50 hPa (~ 20 km ASL). Due to the high horizontal and vertical resolution of the innermost WRF model domains, a fixed numerical time step of 10

seconds was used in domains 3 and 4. Model domains 1 and 2 have larger fixed time steps of 90 and 30 seconds, respectively. We saved WRF model output every 15 minutes to generate a dataset with sufficient temporal resolution to study dynamically evolving atmospheric conditions.

Input Data

North American Model (NAM) 12 km analysis and lateral boundary condition files were used as gridded input data to generate the wind and solar forecasts. NAM data were downloaded from a public data web server[2] maintained by the University Corporation for Atmospheric Research (UCAR). The NAM boundary conditions files are available at 3-hour intervals and were used to update the lateral boundary accordingly throughout the numerical integration. Analysis and 3-hour forecast fields from the Global Forecast System (GFS) model were used for generating the synthetic observations, which are described in greater detail below. GFS data are available at half-degree resolution and were downloaded from a web portal[3] maintained by the National Climatic Data Center (NCDC). Metar, maritime, and mesonet data platform weather observations from the Meteorological Assimilation Data Ingest System (MADIS) [Miller 2005, Miller 2009], were used for model initialization and generating synthetic observations.

Four-Dimensional Data Assimilation

A four-dimensional data assimilation (FDDA) dynamical initialization approach was used for the WRF forecasts over a 6-hour spin-up period prior to the start of the pure forecast. The FDDA approach consists of both an analysis (Stauffer and Seaman, 1994] and observational [Liu et al., 2005] relaxation term that nudges the numerical integration toward the observed state. WRF analysis nudging was used to constrain large-scale atmospheric features based on coarse resolution-gridded analysis fields. Conversely, observational nudging utilizes irregularly spaced weather observations to influence the development of model-simulated small-scale, localized atmospheric phenomena. The objective of using an FDDA based spin-up period is to generate an atmospheric state at the beginning of the simulation that is consistent with available observation data, thus leading to enhanced prediction skill. During the 6-hour spin-up period, WRF analysis nudging was turned on above the boundary for model domains 1 and 2. Analysis nudging was used in the free atmosphere only to allow the model to simulate complex near-surface vertical features that may be absent in the coarser vertical and temporal resolution-gridded analysis fields. Observational nudging of MADIS data was used during the FDDA spin-up initialization on WRF domains 2, 3, and 4. FDDA initialization for WRF forecasting is ideal for this study since it has already been demonstrated to improve short-range wind predictions in the western-central United States (Liu et al., 2009).

Synthetic Observations

Using the FDDA capability described above, we generated synthetic observations of solar flux or hub height wind speed at the renewable resource locations considered in the study and described in detail in Section 2.2. The synthetic observations were used in the absence of onsite meteorological observations because they are the best possible estimate of the true state of the atmosphere that can be achieved with a sophisticated NWP model and data assimilation strategy. As with the forecast runs, a six-hour dynamic initialization spin-up period was used for

---

[2] http://soostrc.comet.ucar.edu/data/grib/nam/
[3] http://nomads.ncdc.noaa.gov/data/gfsanl/

6

generating the synthetic observations. However, unlike the forecast simulations that assimilated observations only during the spin-up period, the model integration for synthetic observations nudged toward analysis fields and assimilated all available weather observations for the entire simulation period. Numerous studies (e.g. Stauffer and Seaman, 1990; Lo et al., 2005; Otte, 2008; Bowden et al., 2012) have demonstrated that FDDA nudging reduces model simulation error and is therefore an effective strategy for generating accurate synthetic observations.

2.2 Renewable Resources

A total of 57 wind and solar resource sites in the Western United States were represented in the study analysis, and their locations are shown in Figure 2. Some large renewable sites outside California were included in the analysis since renewable generation is frequently imported to California from out of state, making their renewable predictability relevant to California grid operations. The study included wind parks, solar thermal / photovoltaic (PV) systems, small solar and distributed solar regions. A total of 14.6 GW of renewable resource installed capacity was included in the study analysis with 10.4 GW coming from wind and 4.2 GW from solar resources. It should be noted that additional renewable resources undoubtedly exist (particularly small solar). However, in the absence of a detailed up-to-date database of installed renewable capacity, only confirmed resources were included in the study. A full list of the renewable resources represented in this study and their associated installed capacity is provided in the Appendix Table 1.
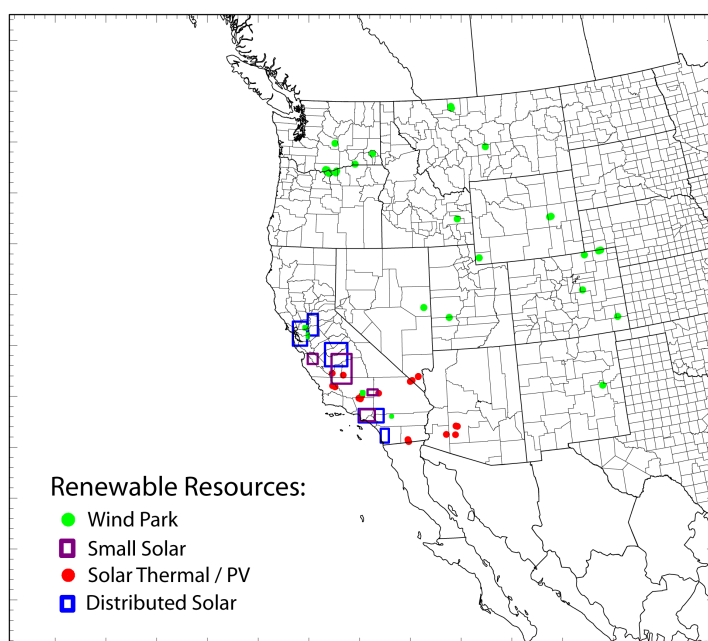
Figure 2. Map showing the location of renewable resources included in the modeling study. The locations of wind parks are denoted by green circles while solar thermal / PV sites are indicated by red circles. Small solar and distributed solar regions are denoted by purple and blue boxes, respectively.

The downward solar flux at solar thermal and PV sites for each model output time was obtained from the single nearest WRF model grid cell. For distributed and small solar regions, an area average solar flux was calculated for each output time from all of the non-water land use grid cells within the regional renewable bounding box.

For wind parks outside of California, the hub height wind speed was calculated from WRF output at a single grid cell by linearly interpolating the vertical wind profile to 80 m. Wind speeds were calculated for 80 m above ground level since this is a common hub height in the wind industry. Data from only a single grid cell was used to calculate the hub height wind speed for wind parks outside of California because they fall within the coarse resolution WRF outer domain. Approximate bounding boxes were used to represent wind parks in the high wind resource regions of California that fall within a WRF 3 km grid spacing domain. An 80 m hub height wind speed was calculated at all grid cells falling within each of the wind park bounding boxes, and then a final wind park average wind speed was calculated for each model output time. The average hub height wind speed for each wind park was ultimately used in the analysis discussed later.

3. Ensemble Member Downscaling

A few hundred thousand functioning model physics configurations are plausible with WRF version 3.5.1. Such a large number of WRF configurations are infeasible to run even with high performance computing resources. In addition, data storage and post-processing at this scale would present their own unique challenges. To reduce the number of combinations used for the ensemble forecast analysis, we employed a strategy of minimizing the redundancy of scheme parameterization approaches and utilized results from a sensitivity test case.

3.1 Sample across sophistication

We developed a preliminary list of physics schemes to include in our study by selecting models that span the level of sophistication and parameterization approaches associated with each of the main WRF physics categories. For example, within the planetary boundary layer (PBL) physics category, some schemes use a simplistic non-local closure scheme to parameterize vertical mixing. Conversely, some PBL schemes are more sophisticated and utilize a prognostic turbulent kinetic energy (TKE) methodology to represent local vertical mixing while other even more advanced schemes predict sub-grid scale TKE terms. Another example is the land surface model (LSM), where the number of soil levels and treatment of vegetation effects on the surface energy balance varies greatly among the available LSM schemes. A list of the physics schemes obtained by sampling among the range of parameterization approaches for each physics category is provided in Table 1. It should be noted that the sensitivity of WRF forecast results to the choice of the surface layer scheme was not studied. Instead, the surface layer scheme recommended by WRF documentation for each PBL scheme was used.

Table 1. Preliminary list of schemes obtained by sampling different model parameterization approaches for each WRF physics category.

Physics category	Scheme			
Planetary Boundary Layer	YSU	MYJ	QNSE	MYNN2
Land Surface Model	Thermal Diffusion	Noah	RUC	Pleim-Xiu
Shortwave Radiation	Dudhia	CAM	RRTMG	FLG
Longwave Radiation	RRTM	RRTMG	FLG	
Cumulus Scheme	Kain-Fritsch	BMJ	Grell-3	Tiedtke
Microphysics	Lin	WSM3	WSM5	

3.2 Latin Hypercube sampling

As discussed in the introduction, we took a sequential approach to determine the smallest number of ensemble members that is as close as possible to the full ensemble in terms of accuracy of its predictions and the magnitude of the ensemble spread. As discussed below, at every stage of the analysis we down-selected to a smaller subset of the initial ensemble and performed a more detailed study of the smaller ensemble than at the previous stage.

Exhaustive sampling of all the combinations of the physics schemes listed in Table 1 would require 4×4×4×3×4×3 = 2304 WRF runs. Since running such a large ensemble would be a sizeable computation exercise even with high performance computing, at the very first stage of the analysis, this 6-dimensional space of physics categories was sampled using the Latin hypercube sampling (LHS) design. LHS is a space-filling design that requires significantly fewer runs than sampling the entire grid of the 6-dimensional space above. It is able to cover the input space efficiently by placing the points in each dimension of the input space so as to ensure that the distribution along each dimension is uniform (see Santner et al., 2003 and Fang et al., 2006 for more details on LHS designs). Figure 3 shows an example of such a design for a 2-dimensional input space.
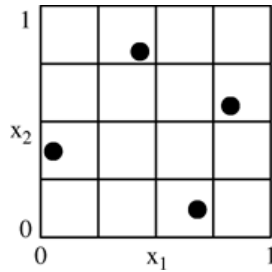

Figure 3. An example of a Latin hypercube sampling design for 2 inputs, x_1 and x_2. Each of the two dimensions in this example is binned into 4 intervals, and each of the 4 bins along each dimension is populated, resulting in 4 sample points. The number of bins along each dimension (in this example and for any number of dimensions) is thus equal to the number of points one wishes to sample.

We created an LHS design with 420 points, or runs, using the R statistical software (R Core Team, 2013) (note that all statistical analysis described in this report was performed using the R software). This number of runs was chosen because it is a multiple of the least common multiple of the unique numbers of levels in each of the 6 dimensions (4×3 = 12 in this case) and was thus a convenient number for the LHS design. In addition, it was in the feasible range of the number of runs we could complete at this stage of the analysis.

A large-scale WRF model physics sensitivity test was run with the 420 ensemble members specified by the LHS over the period of 7 April at 12:00 Z to 9 April at 06:00 Z of 2013. This particular time period was selected due to a large amount of variability in wind speed at the major wind resource locations in California and varying cloud cover over large portions of the WRF model domains.

Upon completion of the WRF sensitivity test, it was revealed that not all of the physics scheme combinations function correctly in WRF at this time. Specifically, WRF runs with the FLG shortwave radiation model that were not coupled with the FLG longwave radiation scheme ran to completion, but produced a zero value for the shortwave flux even during daytime. As a result,

all runs utilizing the FLG shortwave scheme were ultimately discarded. In addition, a total of 6 runs failed for undetermined reasons and could not be successfully completed even with a significant reduction in the numerical time step. Since all these resulted in 111 unusable simulations, a total in 309 completed runs were used for the model analysis. While removing 111 runs from the original design affected the marginal distributions of some of the inputs, thus partly removing the advantages of the LHS design, the impact of this was small. Consequently, we chose to proceed with the 309-run design rather than create a new LHS design and rerun the WRF sensitivity test case.

The 309-member WRF ensemble discussed above was run at each of the 57 renewable generation locations specified earlier. The output of each such member run at each location was a time series of the WRF prediction of either the hub height wind speed (for a wind location) or the shortwave flux (for a solar location) at 15-minute intervals for a period of 42 hours.   Each time series thus consisted of 169 forecasts of the wind speed or the shortwave flux, and we had 309 such time series at each of the 57 locations.

3.3 Forecast Sensitivity Test Results

Our goal was to select a subset of the above ensemble, one that captures as much as possible the spread in WRF predictions of hub height wind speed and shortwave flux at the surface observed in the full ensemble while also retaining the same amount of accuracy as the full ensemble. To do this, we needed to determine the importance of each physics category for the WRF predictions so that schemes that have little or no effect on the predictions can be omitted from further consideration, thus reducing the number of ensemble members.

In order to determine the sensitivities of the predictions to each of the physics categories, we performed an analysis of variance (ANOVA) on the WRF predictions. Among several outputs of such an analysis is the fraction of total variability in the output that can be attributed to each input and any interactions among inputs that are included in the model. Higher fractions imply greater importance of the input (see Neter et al., 1990, pp. 87-98 for details on ANOVA).

Recall that for a given location, the output for each ensemble member is a time series of 169 predictions.   Since these are highly correlated, they cannot be analyzed independently. Consequently, we first performed a principal component analysis (PCA) on the 309 time series in order to obtain 169 uncorrelated linear transformations, or components, of the original output (for details on PCA, see Jolliffe 2002).   Each component consists of 309 values, one for each ensemble member.   Because the components are uncorrelated, they can be analyzed separately.

For a given location, we then performed ANOVA on each of the components, obtaining the fractions of the variance in the component that can be attributed to each physics category and two-way interactions of all the categories. The components themselves have different weights as they account for different proportions of the variance in the output (at most 25 out of 169 components accounted for 99% of the variance in these data across the locations), so we weighed the ANOVA fractions by the fractions of the variance each component accounted for, thus obtaining one weighted average fraction explained by each input or a two-way interaction of inputs at a given location. As an example, Table 2 gives a partial list of the ANOVA results for one location, the Agua Caliente solar project.

Table 2. Fractions of the variance in the output attributed by the ANOVA to each physics category and top 2 category interaction contributors at the Agua Caliente solar project in the 309-member ensemble (the remaining 13 interactions were omitted for brevity). Note that the "x" notation denotes an interaction of two physics categories.

Physics category	Fraction of variance
PBL	0.031
LSM	0.064
SW	0.597
LW	0.061
CU	0.011
MP	0.042
PBL × LSM	0.036
LW × MP	0.059

It should be noted that the use of PCA in the context of ensemble down-selection is not new. For example, Lee et al. (2012) used PCA to analyze the importance of each ensemble member at different locations at a given time point in the forecast trajectory, by examining the amount each ensemble member contributed to the top components. Our use of PCA was slightly different in three ways. First, Lee et al. performed PCA on the forecast errors while we worked with predictions since we did not have the actual observations of either the wind speed or the shortwave flux in this step of the analysis (however, we worked with forecast errors in later stages of the analysis, as discussed in Section 4). Second, we performed the PCA on time series at a given location rather than on locations at a given time point because in our data the temporal correlation was much stronger than the spatial correlation. Third, we used the results of the PCA as inputs to the ANOVA in order to separate the effects of the physics categories while Lee et al. used the PCA directly to determine each specific member's importance without tracking the contribution of each category. Our analysis can thus reveal the importance of each physics category more systematically.

Since ultimately we wanted to obtain one reduced set of ensemble members to use across all locations, we averaged the fractions attributed to each physics category and two-way interaction of categories across locations, weighing each location by its installed capacity relative to the total installed capacity (in MW) across the 57 locations. Note that we did not differentiate between solar and wind locations in this step, but only took into account the fraction of the capacity a given location contributes to the total capacity. Table 3 lists the capacity-weighted average fractions of the variance attributed to each physics category and two-way interactions of categories across the 57 locations. Figure 4 shows a plot of these average fractions (only the top 2 variance-contributing interactions are shown for ease of presentation since the remaining 13 are negligible, as shown in Table 3), as well as the fractions at each location (such as those in Table 2). Moreover, for reference, the plot also shows the capacity-weighted averages across locations when solar and wind locations are considered separately.

Table 3. Capacity-weighted average fractions of the variance in the output attributed by ANOVA to each physics category and all two-way interactions of categories across the 57 locations in the 309-member ensemble. Note that the categories and their interactions are ordered by the fractions, with top contributors listed first, starting with the left-hand column.

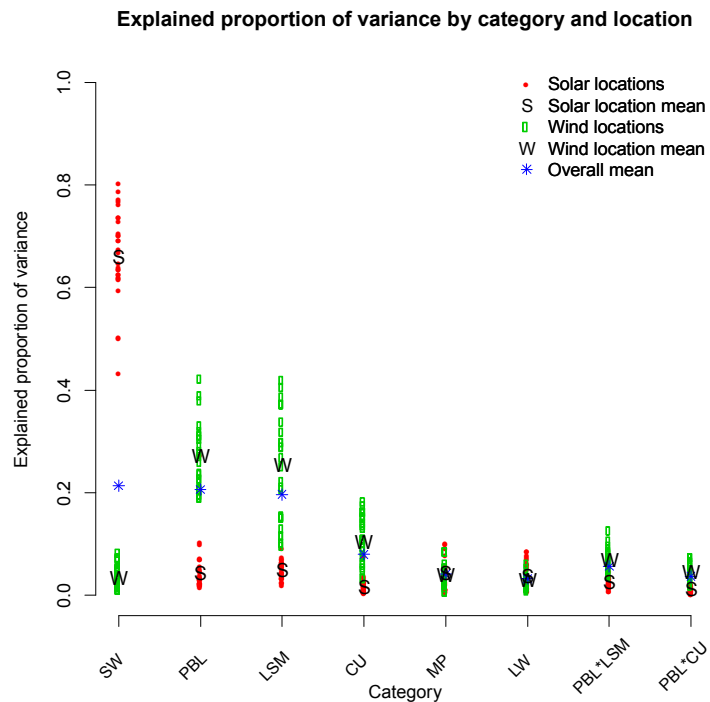| Physics category	Fraction of variance	Physics category	Fraction of variance
SW	0.213	CU × MP	0.012
PBL	0.206	PBL × SW	0.012
LSM	0.196	LSM × MP	0.011
CU	0.079	PBL × LW	0.011
MP	0.041	LSM × SW	0.008
LW	0.032	LSM × LW	0.008
PBL × LSM	0.057	SW × CU	0.007
PBL × CU	0.037	LW × CU	0.007
LSM × CU	0.021	SW × MP	0.007
PBL × MP	0.017	SW × LW	0.004
LW × MP	0.015		

Figure 4. The fraction of the variance attributed to each physics category and top 2 interaction contributors in the 309-member ensemble for each solar and wind location (red and green circles, respectively), their averages across solar and wind locations separately ("S" and "W", respectively) and capacity-weighted averages across all locations put together (blue stars).

As can be seen from both the table and the plot, the shortwave flux (SW), planetary boundary layer (PBL) and land surface model (LSM) physics categories are the top contributors of the variance in the output. The cumulus category (CU) is less important, and the microphysics (MP)

and the longwave (LW) categories have the smallest effect. Moreover, the interactions of the PBL and LSM schemes and PBL and CU schemes are the top contributors among the interactions although their relative importance is much smaller than that of the top 3 main effects.

The plot in Figure 4 also illustrates the impact of averaging across the solar and wind locations. Because the installed capacity across the 57 locations is concentrated quite heavily at the wind locations (10.4 GW out of 14.6 GW, or 71% of the total capacity), the importance of the physics categories to the wind predictions largely determines the capacity-weighted average across all locations. Thus, while PBL and LSM categories played a small role for the solar predictions, since they were important for the wind predictions, they were judged to be important in the overall analysis. At the same time, when a physics category made an overwhelming contribution to solar predictions (such as in the case of SW), that category was still judged important overall despite not playing a large role in the wind predictions.

Based on this ranking, we decided to down-select the schemes of each physics category as follows: retain all levels of SW, reduce the number of PBL and LSM schemes from 4 to 3 each, the number of LW and MP schemes from 3 to 2 each, and the number of CU schemes from 4 to 2. To determine which schemes to retain, we calculated statistical contrasts for all pairs of schemes for all categories except for SW since we had decided not to omit any SW schemes at this stage. Statistical contrasts can be used to estimate the difference between levels of a factor or combinations of factors (see Neter et al., 1990, pp. 576-578 for details on contrasts). The p-value for the test of the hypothesis that a contrast is 0 can be used to judge how significantly different the levels are, with lower p-values implying more significant differences. We obtain a p-value for each contrast for each location. To obtain one overall measure of the difference between each pair of schemes for each physics category across locations, we weighted 1 minus the p-value by the capacity at each location, so the higher the resulting value the greater the difference was judged to be. It should be noted that we only use these values as relative, rather than absolute, measures of the differences between the levels. If one wanted to determine how different all pairs of levels are in an absolute sense, one would have to account for the issue of making multiple comparisons by adjusting the significance level to control the false discovery rate.

The resulting measures of the difference for all the pairs of schemes for each of the 5 relevant physics categories for which the number of schemes were being reduced (i.e., all but SW) are shown in Figure 5. As can be seen from the first plot in the figure, the PBL schemes 2 (MYJ) and 5 (MYNN Level 2.5) are least different when this difference measure was used, suggesting that either of these 2 schemes can be omitted.   Similarly, for the LSM, schemes 3 (RUC) and 7 (Pleim-Xiu) are least different on this measure. Thus, either of these can be omitted. For the CU category, our goal was to omit 2 schemes. The smallest differences were between schemes 1 (KF) and 2 (BMJ), schemes 1 (KF) and 5 (Grell) and schemes 2 (BMJ) and 5 (Grell), suggesting omitting either of these three pairs. The above-mentioned PBL, LSM and CU schemes were therefore all candidates for being omitted. Since the interactions of PBL and LSM and PBL and CU were found to play a relatively important role, we used these to make the final determination regarding the schemes to omit within each category. An analysis of the interaction contrasts led us to drop PBL scheme 2 (MYJ), LSM scheme 3 (RUC) and CU schemes 1 (KF) and 5 (Grell).   The omitted and retained schemes are listed in Table 4.   We thus proceeded to the next step of the analysis with 54 combinations of the retained physics schemes.
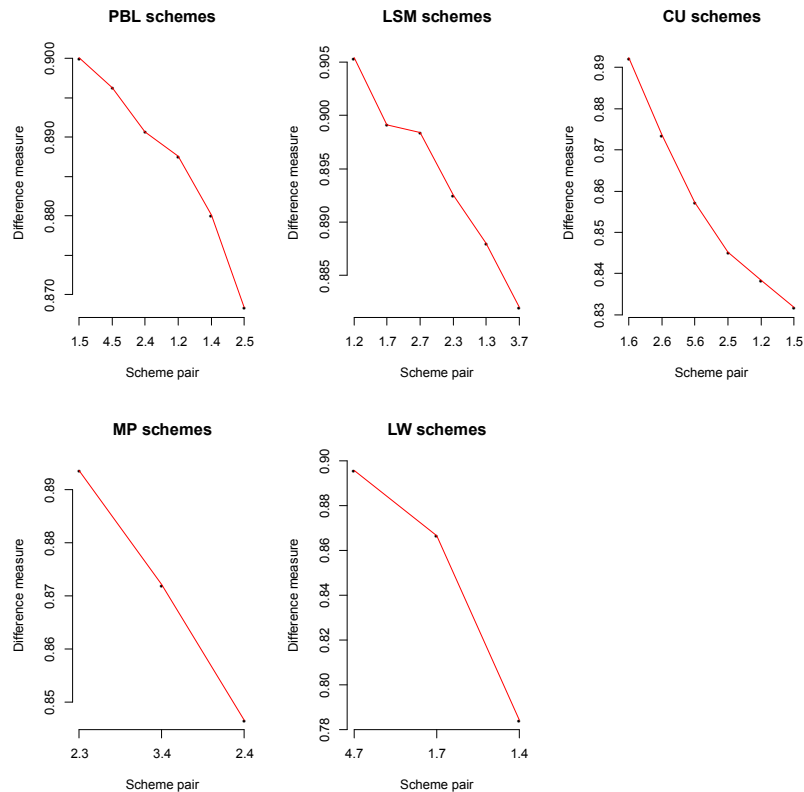
Figure 5. Difference measure for each pair of schemes within each of the 5 physics categories for which the number of schemes were reduced.

Table 4. Omitted and retained schemes within each physics category from the initial 309-member ensemble.

Physics category	Omitted schemes		Retained schemes		
Boundary Layer	MYJ		YSU	QNSE	MYNN2
Land Surface Model	RUC		Thermal Diffusion	Noah	Pleim-Xiu
Shortwave Radiation	*None*		Dudhia	CAM	RRTMG
Longwave Radiation	RRTMG	FLG	RRTM		
Cumulus	Kain-Fritsch	Grell 3D	BMJ	Tiedtke	
Microphysics	Lin	WSM 3-class	WSM 5-class		

4. WRF Renewable Forecast Case Studies

A total of 9 forecast case studies that spanned all seasons of 2013 were performed for the final phase of the research methodology using the derived 54 member ensemble configurations. The case studies were selected based on communication with staff meteorologists at Southern California Edison and San Diego Gas and Electric who suggested the time periods representing synoptic patterns that they associate with dynamically evolving wind and cloud patterns in California. Like the sensitivity test, the case studies were all 42-hour forecasts in duration.

14

4.1 Forecast Verification

Examples of WRF predicted wind and solar forecasts, as well as the synthetically generated observations for the corresponding time periods are provided below. These comparisons demonstrate that the ensemble prediction system generated reasonable wind and solar flux forecast results.

A time series of ensemble-predicted hub height wind speeds at 15-minute intervals at the San Gorgonio wind park for a forecast case study starting at 12:00 Z on 24 April 2013 is shown in Figure 6a. The figure shows the synthetic observation, as well as the minimum, 25th percentile, median, 75th percentile and maximum of the ensemble prediction distribution, at each time point of the forecast period. The ensemble forecast accurately predicted the occurrence of a significant upward wind ramp around forecast hour 11 where the wind speed increased from 5 to 20 m/s. In addition, the vast majority of the synthetic observations fall within the ensemble spread indicating the ensemble design was sufficiently dispersive to capture a dynamic wind change at San Gorgonio. Figure 6b shows the same quantities for another forecast period, starting at 12:00 Z on 20 November 2013. Again, most observations fall within the ensemble spread, and the model was able to accurately forecast hub height wind speeds dropping from 20 m/s at the beginning of the forecast period to as low as 1 m/s at forecast hour 30.
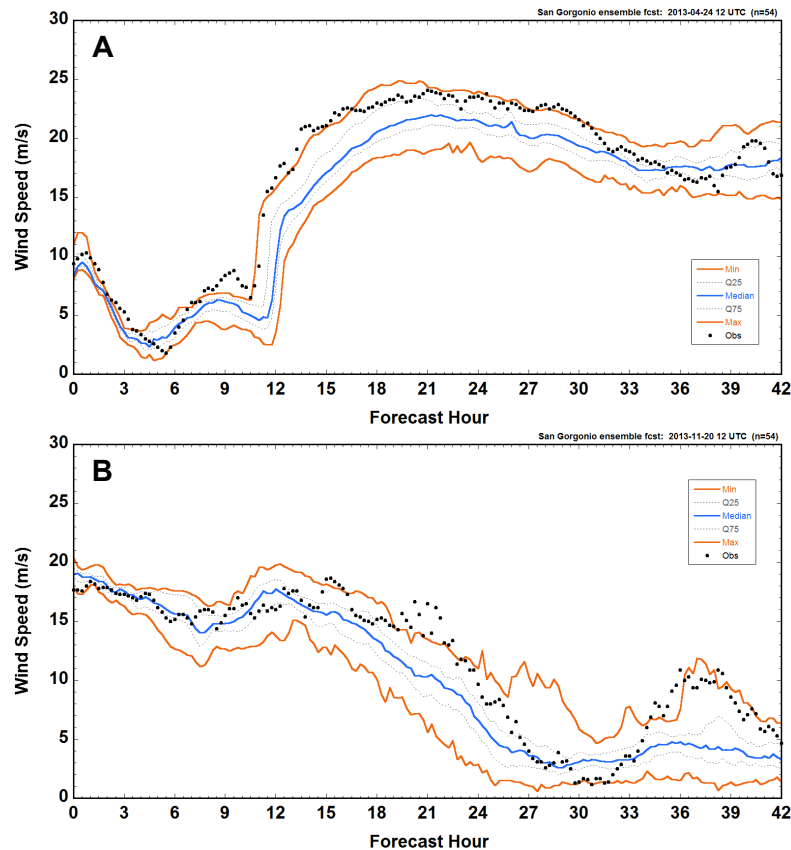

Figure 6. (a) Comparison of hub height synthetic observations (black dots) with ensemble forecast distribution starting at 12:00 Z on 24 April 2013 at the San Gorgonio wind park. (b) Same as (a) but for a forecast period starting at 12:00 Z on 20 November 2013. In both plots, the blue curve indicates the median of the ensemble predictions, the orange curves indicate the minimum and the maximum, and the grey lines indicate the 25th and the 75th percentiles.

Figure 7a shows a typical comparison of synthetic wind speed observations with the ensemble forecast distribution at the Tehachapi wind park for a forecast period starting at 12:00 Z on 24 April 2013. The ensemble forecast captured the major trends of increasing and decreasing wind speeds during the forecast period. However, there were periods during which the synthetic observations fell slightly outside the ensemble spread, indicating that the ensemble was likely under-dispersive at Tehachapi. In general, the ensemble forecasts performed well at Tehachapi, but with less skill than observed at Gorgonio. Figure 7b shows a similar comparison of synthetic wind speed observations with the ensemble forecast distribution at the Altamont wind park for a forecast period starting at 12:00 Z on 6 September 2013. The ensemble accurately predicted the timing and magnitude of wind speed changes. Most synthetic observations fall within the ensemble spread at Altamont, indicating that the forecast design was sufficiently dispersive to generate reliable probabilistic predictions.
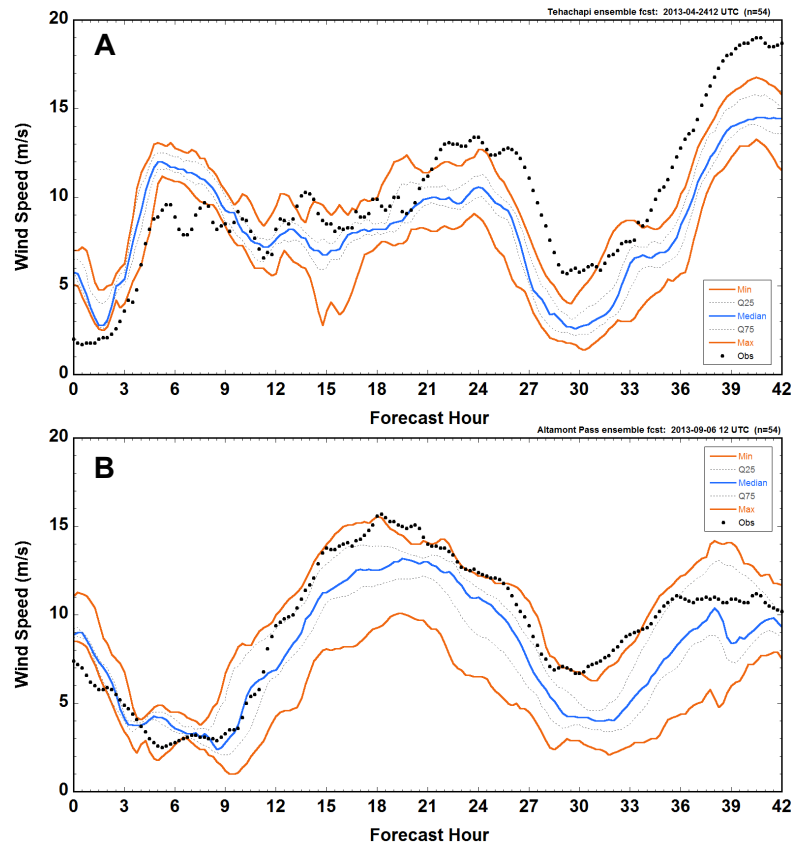

Figure 7. (a) Comparison of hub height synthetic observations (black dots) with ensemble forecast distribution starting at 12:00 Z on 24 April 2013 at the Tehachapi wind park. (b) Same as (a), but at the Altamont Pass wind park and for a forecast period starting at 12:00 Z on 6 September. In both plots, the blue curve indicates the median of the ensemble predictions, the orange curves indicate the minimum and the maximum, and the grey lines indicate the 25th and the 75th percentiles.

A comparison of synthetically generated shortwave flux observations (black dots) to the ensemble forecast distribution for the San Diego distributed solar domain for a 42-hour forecast period starting at 12:00 Z on 5 May 2013 is shown in Figure 8a. Moderate cloud cover was present in the region on the first day of the forecast period, which resulted in ensemble

predicted peak shortwave flux values ranging from 850 to 1000 W/m^2. The peak synthetic observation shortwave flux value on forecast day 1 was around 850 W/ m^2 and fell within the lower limit of the ensemble spread. Widespread cloud cover was observed on the second day of the forecast period with a peak shortwave flux observation around 400 W/m^2. The ensemble spread accurately captured the significant reduction in shortwave flux relative to day one, though the observation was on the lower end of the forecast distribution. It is also interesting to note the enormous spread in predicted peak shortwave flux on forecast day 2 with values ranging from 375 to 950 W/m^2. The magnitude of the ensemble spread highlights the degree of uncertainty associated with solar forecasting in the San Diego area.

Synthetic shortwave flux observations and ensemble forecast distribution for the same time period as Figure 8a, but for the San Francisco Bay area distributed solar domain, are shown in Figure 8b. As in the case of the San Diego domain, there was increasing cloud cover from forecast day 1 to 2 with peak synthetic shortwave flux observations decreasing from around 800 to 600 W/m^2. All of the synthetic observations fall within the ensemble spread and the median of the forecast distribution (blue line) matches the observations well. This excellent agreement between observation and prediction is typical of the WRF case study results for the San Francisco Bay region.
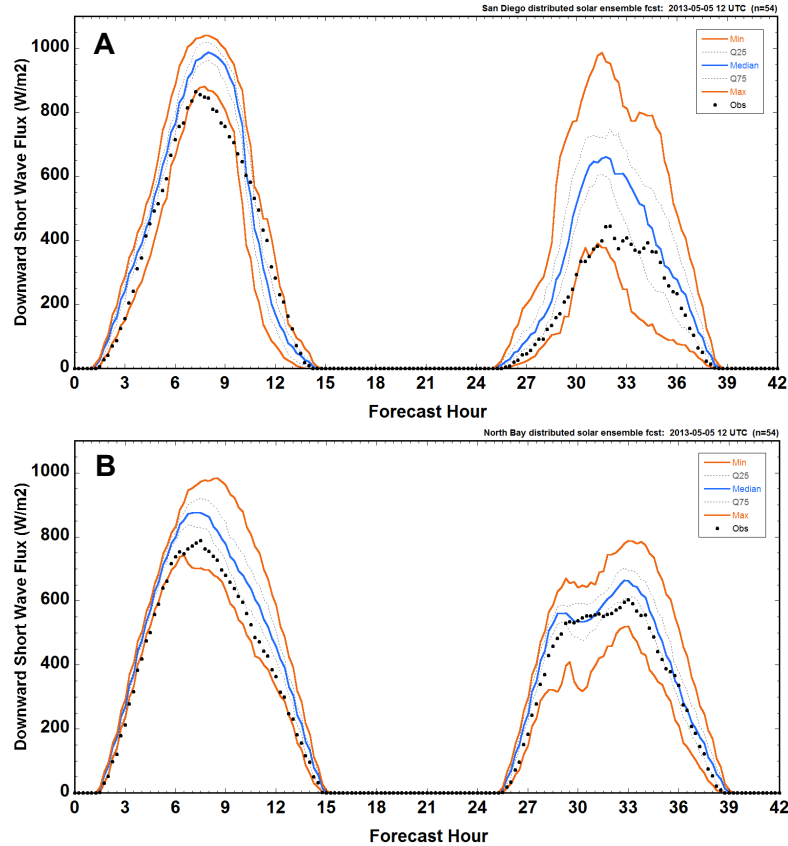

Figure 8. Comparison of shortwave flux at the surface synthetic observations (black dots) with ensemble forecast distribution starting at 12:00 Z on 5 May 2013 for the San Diego distributed solar domain. (b) Same as (a), but for the San Francisco Bay area distributed solar domain. In both plots, the blue curve indicates the median of the ensemble predictions, the orange curves indicate the minimum and the maximum, and the grey lines indicate the 25th and the 75th percentiles.

4.2 Case Study Forecast Analysis

The objective of the final stage of the analysis was to further reduce the 54-member ensemble to a more efficient subset. Since for each of the case studies we also have synthetic observations, we could compute forecast errors at every time point for each location and case study and, consequently, the root mean squared error (RMSE) of each ensemble member, given by

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^{n} e_t^2},$$

where e_t is the forecast error at time point t. We obtained the capacity-weighted average RMSE for all wind and solar locations separately, for each ensemble member and case study, and these are shown in Figure 9. As Figure 9a shows, the RMSE values at the solar locations for case study 9 were markedly lower for all ensemble members than for all other case studies. Figure 9b shows that the same was true, albeit to a lesser extent, for case study 8 at the wind locations. This is because wind speeds turned out to be low at the major wind resource locations of Tehachapi and Gorgonio, and cloud cover extent was minimal for these two case studies. As a result, these two case studies were removed from further consideration so that they would not bias the results. Thus, in the remainder of this report we discuss the results for case studies 1-7 only.
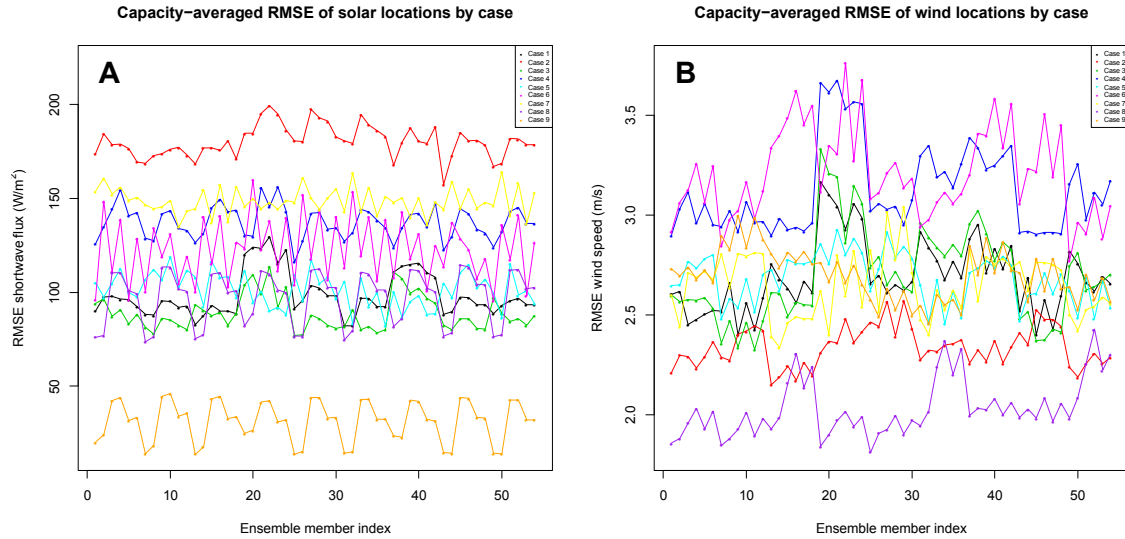

Figure 9. The capacity-weighted average RMSE of (a) solar and (b) wind locations by case (line color) and ensemble member.

Since, as mentioned earlier, we had access to synthetic observations of the wind speed or the shortwave flux for each location and case study, we repeated an equivalent analysis described previously for the 309-member ensemble, but now applying it to forecast errors rather than predictions. Note that since only 1 scheme was retained for each of the LW and MP categories, there were now only 4 physics categories with varying schemes.

As in the case of the 309-member ensemble, we began with an ANOVA for each case study. This yielded fractions of the variance in the forecast errors in the 54-member ensemble explained by

the differences in the schemes within each of the four physics categories for each of the case studies. These are listed in Table 5 and shown in Figure 10. The latter shows that the physics categories' contributions to the overall forecast error variance tended to be consistent across cases, that is, no category was important for some cases, but not for others.

Table 5. Capacity-weighted average fractions of the variance in the forecast errors attributed by ANOVA to each physics category and two-way interactions of categories across the 57 locations in the 54-member ensemble, for each of case studies 1—7. Also listed in the last column are the averages of these fractions across the 7 case studies. Note that the physics categories and their interactions are ordered by the average fraction, with top contributors listed first.

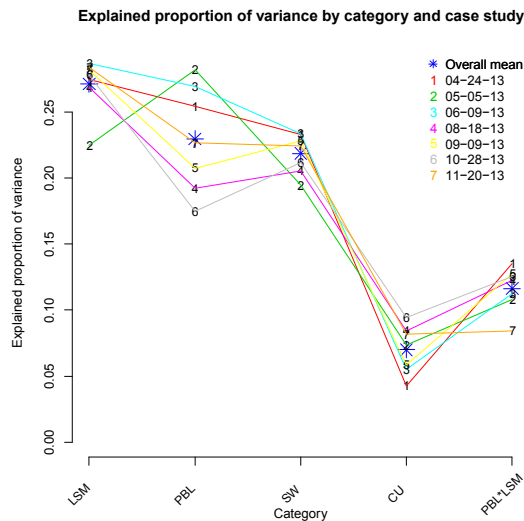| Physics category | Case study | | | | | | | Average |
	1	2	3	4	5	6	7	
LSM	0.275	0.225	0.287	0.269	0.282	0.278	0.284	0.271
PBL	0.254	0.282	0.269	0.192	0.207	0.175	0.226	0.230
SW	0.233	0.195	0.233	0.206	0.228	0.212	0.224	0.219
CU	0.042	0.074	0.055	0.084	0.058	0.094	0.082	0.070
PBL × LSM	0.135	0.108	0.112	0.123	0.127	0.125	0.084	0.116
PBL ×CU	0.016	0.036	0.015	0.030	0.028	0.047	0.027	0.028
LSM×CU	0.019	0.023	0.014	0.033	0.028	0.032	0.019	0.024
LSM×SW	0.013	0.020	0.007	0.026	0.017	0.015	0.021	0.017
PBL×SW	0.010	0.026	0.005	0.022	0.014	0.014	0.021	0.016
SW×CU	0.003	0.013	0.002	0.015	0.009	0.007	0.011	0.009

Figure 10. The fraction of the variance attributed to each physics category and top category interaction contributor in the 54-member ensemble for each of case studies 1-7 (labeled with these numbers in the plot). The lines are drawn for reference to help follow each case's fractions.

To make an overall assessment of the categories' contributions across the seven case studies, the seven fractions for each physics category and interaction were averaged, weighing the case studies equally. These average fractions are listed in the last column of Table 5 and shown as blue stars in Figures 10 and 11. They led to very similar conclusions as those reached in the 309-member ensemble analysis: the LSM, PBL and SW physics were of highest and approximately

equal importance, while the CU physics was relatively less important compared to the other three. Moreover, the interaction of the PBL and the LSM physics was still important.
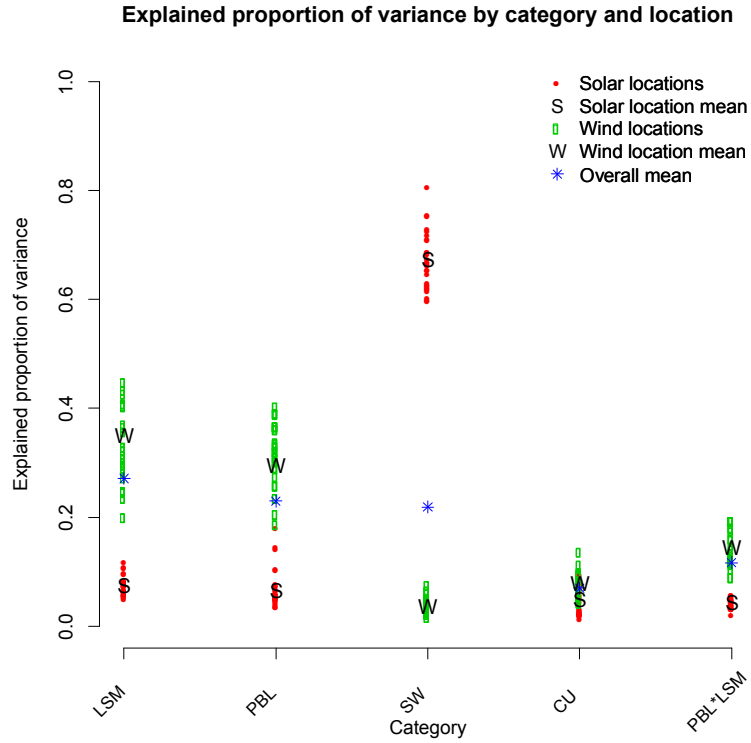

Figure 11. The fraction of the variance attributed to each physics category and top interaction contributor in the 54-member ensemble for each solar and wind location (red and green circles, respectively), their averages across solar and wind locations separately ("S" and "W", respectively) and capacity-weighted averages across all locations put together (blue stars).

Since, according to the ANOVA, the CU physics category did not have a large impact on the forecast errors, one of the two CU schemes was omitted. To determine whether any of the schemes within each of the other 3 physics categories could be omitted, we used statistical contrasts in the same way as described in Section 3.3 for the 309-member ensemble. Figure 12 shows the measure of the difference between all pairs of schemes for each the 3 most important categories – PBL, LSM and SW – by case study. The plots in the figure also show the means of these difference measures for each pair of schemes across the 7 case studies. All pairs of schemes within the PBL and LSM categories have similar difference measures. For the SW category, however, the difference measure for schemes 3 (CAM) and 4 (RRTMG) is markedly lower than for the other two pairs of schemes. This suggests that either of these SW schemes can be omitted. Dropping one of the CU and one of the SW schemes will result in 18 members, which is a highly feasible ensemble size to run for real-time ensemble weather forecasting.
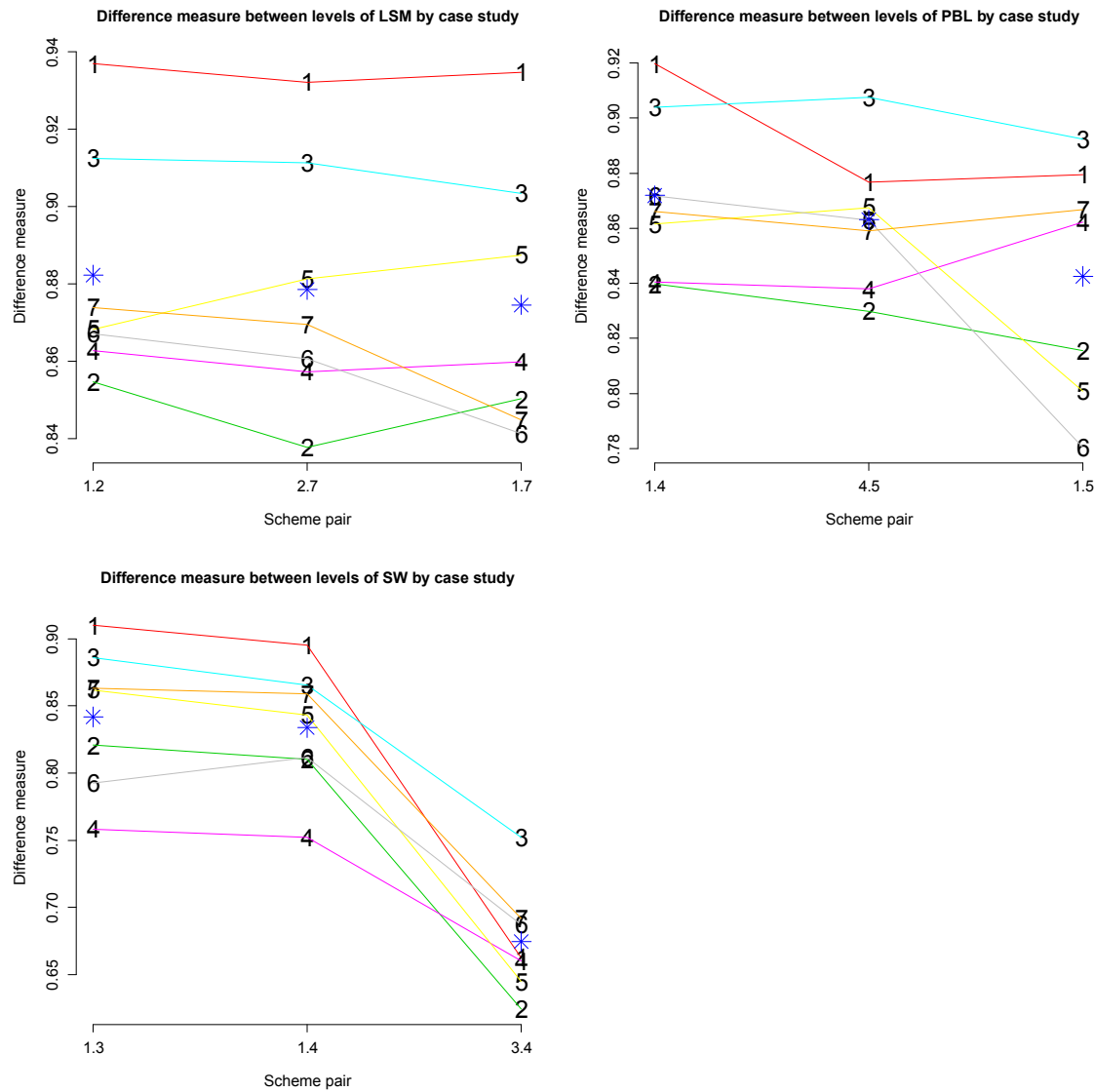
Figure 12. Difference measure for each pair of schemes for each of the 3 top physics category contributors, by case study (labeled with their number in the plots). Also shown are the overall means of the difference measure for each pair of schemes, averaged across the case studies. All the symbols and the colors of the lines are as in Figure 10.

All that remained to finalize the 18-member ensemble was to determine which CU scheme (2 or 6) and SW scheme (3 or 4) to retain. Figure 13 shows the plot of the mean RMSE values (averaged over case studies and location, with capacity weighting for locations) of the ensemble members with each combination of these schemes. Also shown are the grand means of the RMSE, i.e., averages of the mean RMSEs over all the ensemble members for each combination. The plot shows that the combinations with CU = 2 (BMJ) tend to have slightly higher RMSEs than those with CU = 6 (Tiedtke), suggesting that CU = 2 should be omitted. There is no discernible difference between the RMSEs of ensemble members with SW = 3 and SW = 4.
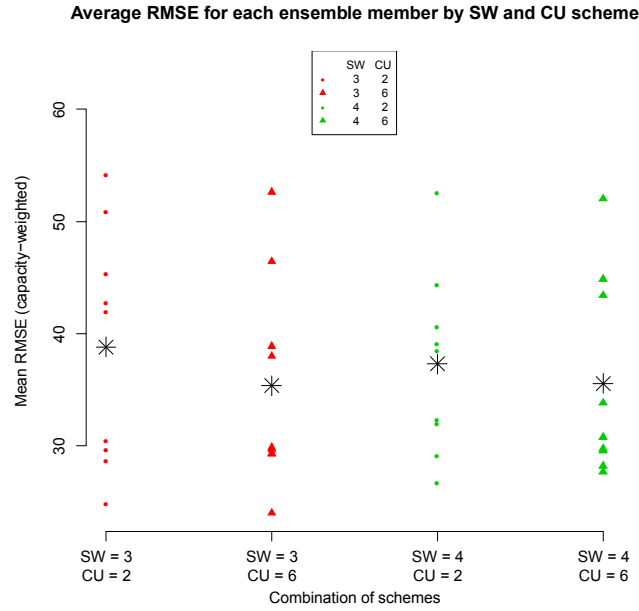
21

Average RMSE for each ensemble member by SW and CU scheme

Figure 13. The mean ensemble RMSE (capacity-weighted average over location and average over case studies for each member) by the SW (2 = red, 3 = green) and CU scheme (2 = circle, 6 = triangle) combination. There are 9 points for each combination of the two categories, one for each ensemble member with that combination. The mean for each SW and CU scheme combination is marked with a black star. The values for SW = 1 are not shown since this scheme was not considered for omission.

To decide which of the SW schemes to omit, we also considered the amount of spread in the reduced ensemble relative to the full ensemble since an ideal reduced ensemble would have spread close to that of the original ensemble. For each location and case study, at each time point, we calculated the ratio of the standard deviation in the reduced 18-member ensemble to that of the original 54-member ensemble. These ratios were averaged across all the 169 time points, resulting in one mean ratio value for each location and case study. The mean ratios were in turn averaged over locations, weighing locations by their relative capacities. We obtained 7 such capacity-weighted ratio averages, one for each case study. A ratio of 1, less than 1 and greater than 1 indicates that on average the reduced ensemble's spread is the same, smaller or greater than that of the original one, respectively. Thus, an average ratio equal to 1 is ideal.

These 7 ratio averages are plotted in Figure 14 for the reduced ensembles, with each reduced ensemble obtained by dropping one of the four combinations of the SW and CU schemes considered for omission (expressed in terms of the retained schemes on the x-axis of the plot). The figure indicates that ensembles obtained by omitting SW = 4 (i.e., retaining SW = 1 and 3) have average ratios closer to 1 than those obtained by omitting SW = 3 (i.e., retaining 1 and 4), suggesting that SW = 4 (RRTMG) should be omitted. Indeed, retaining SW = 1 and 3 (Dudhia and RRTMG) and CU = 6 (Tiedtke) (as suggested by the RMSE comparison in Figure 14), results in an average spread ratio closest to 1 of the 4 combinations.
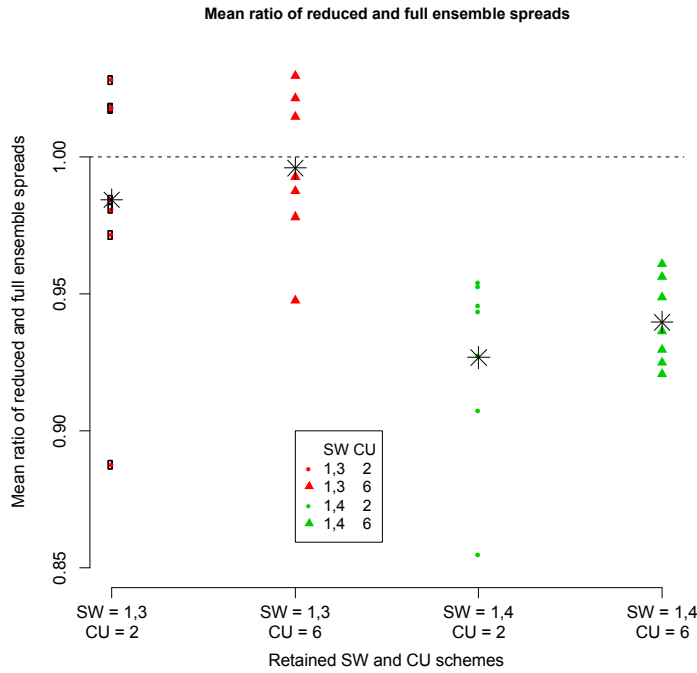
Figure 14. Mean ratio of the reduced 18-member ensemble spread to that of the full 54-member ensemble (ratio equal to, less than or greater than 1 implies the same, smaller and greater average spread, respectively, in the reduced ensemble relative to that in the full one) for each of the reduced ensembles (obtained by retaining only the SW and CU schemes indicated on the x-axis; colors and symbols distinguish the four combinations of the retained schemes). The stars indicate the means of these ratios for each of the ensembles, while the dashed line indicates the average ratio of 1 (equal spread in the reduced and full ensembles on the average) for reference. See the text for details on how the mean ratios were calculated.

To ensure that the 18-member ensemble resulting from omitting SW = 4 and CU = 2 schemes has the same level of accuracy as the original 54-member ensemble, we compared the RMSEs of the two ensembles. The distributions of the mean RMSEs (averaged across locations using capacity weighting) by case study are shown side by side in the form of boxplots for the original and the reduced ensembles at solar and wind locations in Figures 15a and 15b, respectively. Also shown in both figures are the averages of these across the 7 case studies. The two distributions are mostly similar when considered individually for each case study and when the RMSEs are averaged across the case studies. The RMSE values tend to be higher for case studies 6 and 7 at the solar locations, but even in these cases, the two distributions overlap (and in some cases, such as case study 7 at wind locations, it is the reduced ensemble that tends to yield lower RMSEs). Thus, we can conclude that the accuracy of the reduced 18-member ensemble is on par with that of the original 54-member ensemble.
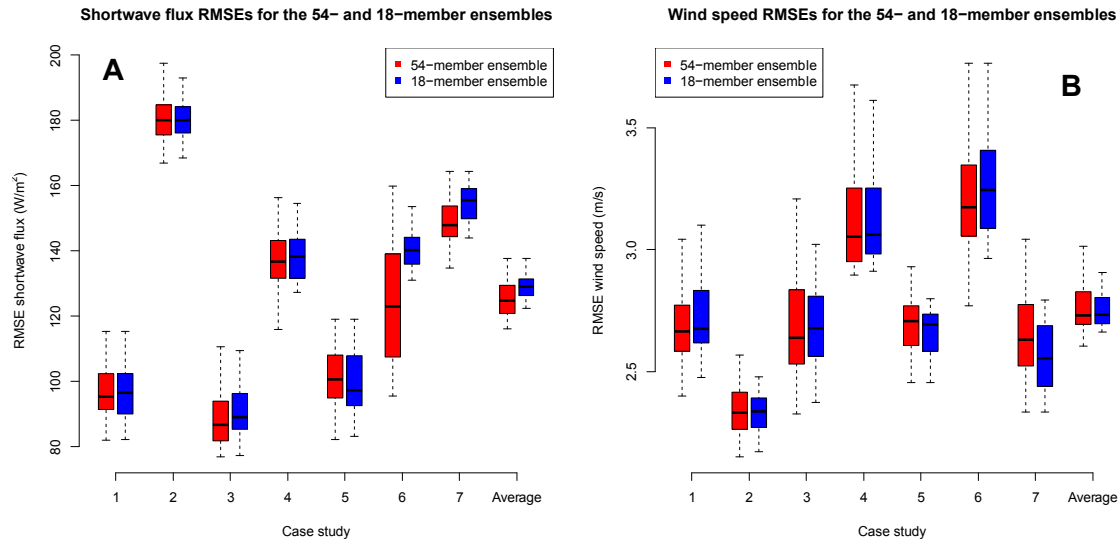
Figure 15. Distributions of RMSEs of the original 54-member (in red) and the reduced 18-member (in blue) ensembles at (a) solar and (b) wind locations, separately by case study, and averaged across the case studies.

Based on results of the WRF forecast error analysis, we thus make the following recommendation for the physics schemes that should be included in the 18-member multi-physics ensemble for predicting renewable generation for current California wind and solar resources:

Physics category	Retained schemes		
Boundary Layer	YSU	QNSE	MYNN2
Land Surface Model	Thermal Diffusion	Noah	Pleim-Xiu
Shortwave Radiation	Dudhia	CAM	
Longwave Radiation	RRTM		
Cumulus	Tiedtke		
Microphysics	WSM 5-class		

5. Conclusions

Our analysis of the WRF ensemble forecasts of wind and solar resources leads to the following conclusions:

- the WRF physics categories that were the primary contributors to ensemble wind speed prediction and forecast error variability were the planetary boundary layer and land surface models

- to a lesser degree, the interaction of the boundary layer and land surface schemes contributed to ensemble wind speed prediction and forecast error variability and should be represented in an ensemble design

24

- the shortwave radiation model was the dominant source of shortwave flux prediction and forecast error variability for California solar renewable resources

- the cumulus, longwave radiation, and microphysics categories were found to have minimal impact on wind and solar predictions in California

- analysis of variance of a limited number of forecast case studies indicated that only 18 ensemble members were needed to generate accurate renewable forecasts with sufficient spread to capture dynamically evolving atmospheric conditions that affect wind and solar generation in California

6. Future Work

Several caveats of the above analysis should be noted. First, in the analysis of the WRF forecast errors, synthetic rather than real observations were used. While synthetic observations are the best possible substitute for the actual ones, using the former in place of the latter means that the forecast error values have some amount of error in them. The conclusions from our analysis need to be tested with real observations in future studies.

Second, we used what is perhaps the most obvious approach for combining the ANOVA results across the 57 locations. However, more optimal ways for doing this may exist, and we did not explore other options in this project due to time constraints. Such an exploration should certainly be a part of any future work in this area. The same is true for creating the difference measure to judge the degree to which any pair of schemes within a category is similar to one another. Similarly, more sophisticated ways to obtain an overall measure of a spread of an ensemble than discussed in this work should be explored.

In this work, we weighed the case studies equally. A better weighting of the case studies may be by their ramp characteristics, such as the presence of ramps (sudden changes in generated power), ramp magnitude and duration. Since larger and longer ramp events present a challenge to utilities and grid operators, it may be useful to emphasize case studies with such events in determining the optimal ensemble. While ramps could be defined in terms of the renewable resource, such as wind speed or the shortwave flux, ideally this would be done in terms of generated power. Obtaining such data may be difficult, but it would allow for a much more direct assessment of the quality of a particular ensemble member in general and in terms of ramp prediction in particular.

In addition, the conclusions of this research are based on a limited number of forecast case studies. Though the individual study forecasts span all seasons, additional case studies need to be run to ensure the findings of this study are valid for a broader range of atmospheric conditions.

References

Berner, J., S. Y. Ha, J. P. Hacker, A. Fournier, C. Snyder, 2011. Model Uncertainty in a Mesoscale Ensemble Prediction System: Stochastic versus Multiphysics Representations, *Monthly Weather Review* 139, pp. 1972–1995.

Bowden, Jared H., Tanya L. Otte, Christopher G. Nolte, Martin J. Otte, 2012. Examining Interior Grid Nudging Techniques Using Two-Way Nesting in the WRF Model for Regional Climate Modeling, *Journal of Climate*, 25, pp. 2805–2823.

Eckel, F. A. and C. F. Mass, 2005. Aspects of Effective Mesoscale, Short-range, Ensemble Forecasting, *Weather Forecasting 20*, pp. 328–350.

Edmunds, Thomas, Alan Lamont, Vera Bulaevskaya, Carol Meyers, Jeffrey Mirocha, Andrea Schmidt, Matthew Simpson, Steven Smith, Pedro Sotorrio, Philip Top, and Yiming Yao Lawrence Livermore National Laboratory, 2013. *The Value of Energy Storage and Demand Response for Renewable Integration in California.* California Energy Commission. Publication number: CEC-500-10-051.

Fang, K.-T., R. Li, and A. Sudjianto*,* 2006. *Design and Modeling for Computer Experiments* (Chapman & Hall), pp. 47-51.

Jolliffe, I.T. , 2002. *Principal Component Analysis*, 2nd edition (Springer).

Hacker, J. P., S. Y. Ha, C. Snyder, J. Berner, F. A. Eckel, E. Kuchera, M. Pocernich, S. Rugg, J. Schramm, and X. Wang, 2011. The U.S. Air Force Weather Agency's Mesoscale Ensemble: Scientific Description and Performance Results. *Tellus 63A*, pp. 625-641.

Hou, D., E. Kalnay, and K. Drogemeier, 2001. Objective Verification of the SAMEX '98 Ensemble Experiments, *Monthly Weather Review 129*, 73–91.

Lee, J.A., W.C. Kolczynski, T.C. McCandless, and S.E. Haupt, 2012. "An Objective Methodology for Configuring and Down-Selecting an NWP Ensemble for Low-Level Widn Prediction." *Monthly Weather Review* 140, 2270-2286.

Lo, J. C. F., Z. L. Yang, and R. A. Pielke Sr., 2008. Assessment of Three Dynamical Climate Downscaling Methods Using the Weather Research and Forecasting (WRF) Model, *Journal of Geophysical Research 113*, D09112, doi:10.1029/2007JD009216.

Liu, Y., A. Bourgeois, T. Warner, S. Swerdlin and J. Hacker, 2005. An Implementation of Obs-nudging-based FDDA into WRF for Supporting ATEC Test Operations, *2005 WRF User Workshop*, Paper 10.7.

Liu, Y., T. Warner, W. Wu, G. Roux, W. Cheng, Y. Liu, F. Chen, L. Delle Monache, W. Mahoney and S. Swerdlin, 2009. A versatile WRF and MM5-based weather analysis and forecasting system for supporting wind energy prediction. 23rd WAF / 19th NWP Conf. , AMS, Omaha, NE. June 1-5, 2009. Paper 17B.3

Miller, P.A., M. Barth, L. Benjamin, D. Helms, M. Campbell, J. Facundo, and J. O'Sullivan, 2005. The Meteorological Assimilation and Data Ingest System (MADIS): Providing Value-added Observations to the Meteorological Community, *21st Conference on Weather Analysis and Forecasting*, Washington, D.C., American Meteorological Society, 2005.

Miller, P.A., M.F. Barth, L.A. Benjamin, 2009. A 2009 Update on the NOAA Meteorological Assimilation Data Ingest System (MADIS), *25th Conference on International Interactive Information and Processing Systems (IIPS) for Meteorology, Oceanography, and Hydrology*, Pheniz, AZ, American Meteorological Society, 2009.

Murphy, J., D. Sexton, D. Barnett, G. Jones, M. Webb, M. Collins, and D. Stainforth, 2004. Quantification of Modeling Uncertainties in a Large Ensemble of Climate Change Simulations, *Nature 430*, pp. 768–772.

Neter, J., W. Wasserman, and M.H. Kutner, 1990. *Applied Linear Statistical Models*, 3rd edition (Irwin).

Otte, Tanya L., 2008. The Impact of Nudging in the Meteorological Model for Retrospective Air Quality Simulations, Part I: Evaluation Against National Observation Networks, *Journal of Applied Meteorology and Climatology 47*, pp. 1853–1867.

R Core Team, R, 2013. *A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing (Vienna, Austria).

Santner,T.J., B.J. Williams, and W.I. Notz, 2003. *The Design and Analysis of Computer Experiments* (Springer), pp. 127-132.

Skamarock, W. C., J. B. Klemp, J. Dudhia, D. O. Gill, D. M. Barker, M. G. Duda, X-Y. Huang, W. Wang and J. G. Powers, 2008. A Description of the Advanced Research WRF Version 3. NCARTech Note, NCAR/TN–475+STR, 113 pp.

Stauffer, D. R., and N. L. Seaman, 1990. Use of Four-dimensional Data Assimilation in a Limited-area Mesoscale Model, Part I: Experiments with Synoptic-scale Data, *Monthly Weather Review 118*, pp. 1250-1277.

Stauffer, D. R., and N.L. Seaman, 1994. On Multi-Scale Four-Dimensional Data Assimilation, *Journal of Applied Meteorology 33*, pp. 416-434.

Stensrud, D., J. W. Bao, and T. T. Warner, 2000. Using Initial Condition and Model Physics Perturbations in Short-range Ensemble Simulations of Mesoscale Convective Systems, *Monthly Weather Review 128*, pp. 2077–2107.

Acknowledgments

Appendix Table 1: List of wind and solar renewable resource sites and their associated installed capacities that were included in the study.

Station Name	Renewable Type	Installed Capacity (MW)	State
Tehachapi_Alta_Wind_Energy_Center	Wind	1710	CA
San_Gorgonio_Pass_Wind_Farm	Wind	619	CA
Altamont_Pass_High_Winds_Wind_Farm	Wind	768	CA
Shiloh_Wind_Farm	Wind	300	CA
Cedar_Creek_Wind_Farm	Wind	550	CO
Peetz_Wind_Farm	Wind	400	CO
Cedar_Point_Wind_Farm	Wind	243	CO
Northeastern_Colorado_Wind_Energy_Center	Wind	174	CO
Colorado_Green_Wind_Farm	Wind	162	CO
Goshen_Wind_Farm_II	Wind	125	ID
Glacier_Wind_Farm	Wind	210	MT
Rim_Rock_Wind_Farm	Wind	189	MT
Judith_Gap_Wind_Farm	Wind	135	MT
Spring_Valley_Wind_Farm	Wind	152	NV
New_Mexico_Wind_Energy_Center	Wind	204	NM
Shepherds_Flat_Wind_Farm	Wind	845	OR
Biglow_Canyon_Wind_Farm	Wind	450	OR
Klondike_Wind_Farm	Wind	400	OR
Stateline_Wind_Project	Wind	300	OR
Leaning_Juniper_Wind_Project	Wind	201	OR
Milford_Wind_Corridor_Project	Wind	305	UT
Windy_Point_Windy_Flats	Wind	400	WA
Lower_Snake_River_Wind_Project	Wind	343	WA
Wild_Horse_Wind_Farm	Wind	229	WA
White_Creek_Wind_Power_Project	Wind	204	WA
Big_Horn_Wind_Farm	Wind	200	WA
Glenrock_wind_farm	Wind	237	WY
Top_of_the_World	Wind	200	WY
Wyoming_Wind_Energy_Center	Wind	144	WY
Ivanpah_Solar_Power_Facility	Solar Thermal	392	CA
Solar_Energy_Generating_Systems	Solar Thermal	354	CA
Solana_Generating_Station	Solar Thermal	280	AZ
Nevada_Solar_One	Solar Thermal	64	NV
Agua_Caliente_Solar_Project	Solar PV	251	AZ
Topaz_Solar_Farm	Solar PV	300	CA
California_Valley_Solar_Ranch	Solar PV	250	CA
Imperial_Solar_Energy_Center_South	Solar PV	200	CA
Copper_Mountain_Solar_Facility	Solar PV	150	NV

Mesquite_Solar_project	Solar PV	150	AZ
Catalina_Solar_Project	Solar PV	143	CA
Campo_Verde_Solar_Project	Solar PV	139	CA
Centinela_Solar_Energy_Project	Solar PV	125	CA
Arlington_Valley_Solar_II	Solar PV	125	AZ
Antelope_Valley_Solar_Ranch	Solar PV	100	CA
Alpine_Solar	Solar PV	66	CA
Alpaugh_Solar_Plant	Solar PV	60	CA
Silver_State_North	Solar PV	60	NV
Avenal_Solar_Facility	Solar PV	58	CA
Dist_Solar_1_Central_Valley_Sacramento	Solar Dist.	111	CA
Dist_Solar_2_Central_Valley_Fresno	Solar Dist.	111	CA
Dist_Solar_3_North_Coast_Bay_Area	Solar Dist.	111	CA
Dist_Solar_4_South_Coast_Los_Angeles	Solar Dist.	111	CA
Dist_Solar_5_South_Coast_San_Diego	Solar Dist.	111	CA